J-025

ＴＶ視聴者の顔向きと表情に基づく番組評価推定

The rating estimation of TV programs based on viewer's head pose and facial expressions

高橋 正樹 Masaki Takahashi　奥田 誠 Makoto Okuda　Simon Clippingdale Simon Clippingdale　山内 結子 Yuko Yamanouchi　苗村 昌秀 Masahide Naemura

1. まえがき

ユーザ個人に最適化された個人化サービスの実現に向けては，各人の嗜好や関心を理解する必要がある．今回，TV の視聴状況を計測することで，番組内容に応じて時事刻々変化するユーザの興味・関心を自動推定するシステムを開発した．家庭用映像センサで観測したデータからユーザの顔向きと表情を計測し，得られた結果から各番組への評価を推定する．顔向きと表情，それぞれ複数のモジュールで並列処理することで，実生活空間でも利用可能な頑健なシステムとした．家庭を模した環境で実験を行い，提案システムが視聴者の趣味・嗜好の理解に有効であることを確認した．

2. 視聴状況推定システム

2.1 概要

視聴状況推定システムは，1 台の家庭用映像センサ（Microsoft Kinect [1]）から得られる情報から，TV 視聴者の「顔向き」と「表情」を計測する．処理の流れを図 1 に示す．カメラ画像を入力とした 2 次元モジュール群，カメラ画像と奥行き画像を入力とした 3 次元モジュール群，および注視状態と表情表出それぞれの検出器（2 値 SVM）で構成される．システムからは注視状態（TV を見ているか否か）と表情表出（表情が出ているか否か）の推定結果がリアルタイム出力される．

各モジュール群それぞれに顔検出・追跡処理，顔向き推定処理，および表情推定処理が行われる．2 つのモジュールが互いにパラメータを他方に供給しながら相補的に視聴状態を推定している点が本システムの特長であり，この構成が，ユーザの自由な振る舞いに対する頑健性を高めている．各番組視聴後，視聴時間内における注視状態の時間率および表情表出頻度を指標とし，番組への評価を自動推定する．

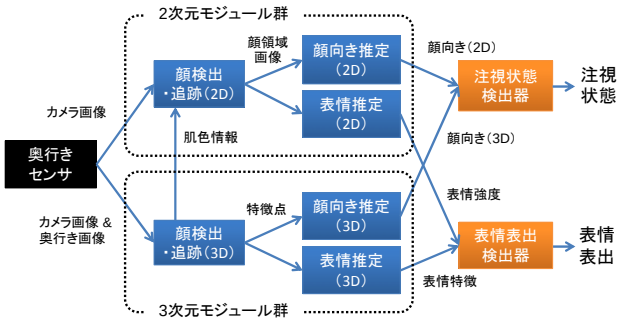

図 1　視聴状況推定の流れ

NHK 放送技術研究所, NHK STRL

2.2 2 次元モジュール群

・　顔検出・追跡、顔向き推定モジュール

2 次元顔向き推定モジュールは，ユーザを撮影した映像から顔領域を検出・追跡し，顔の向きをリアルタイムで計測する[2].

本モジュールは，まず指定された肌色情報（YCbCr 色空間での Cb, Cr）を用いて肌色領域を検出し，領域内からユーザの顔を検出する．その後，各フレームで可変テンプレートを算出し，あらかじめ登録された様々な顔向きの可変テンプレートとのマッチングにより，ユーザの顔向きを推定する．

可変テンプレートは，9 点の特徴点の位置（x-y 座標）と，各特徴点で計測されたガボールウエーブレット特徴からなる．マッチングでは，まず検出された正面顔の可変テンプレートと登録済みの人物不特定可変テンプレートとの照合が行われる．以降のフレームでは，追跡状況に応じて照合する顔向きを変化させながら頑健なマッチング処理を施し，上下・左右方向の顔向きを推定する。

本モジュールでは肌色情報の初期化・更新が必要となるが，3 次元モジュールからパラメータを取得することで，頑健な顔検出・追跡処理を実現している．

・　表情推定モジュール

2 次元表情推定モジュールは，TV 視聴者の顔表情（怒り，嫌悪，恐れ，喜び，悲しみ，驚きの 6 分類）を認識し、その強度を推定する[3].

まず顔領域内のローカルバイナリパターンを計算し，得られた特徴量を事前に求めた回帰式に当てはめ，表情それぞれの生起確率，およびその強度を計算する．これに並行し，単純ベイズ確率モデルによる表情の推定も行う．最後に，回帰式による結果とベイズ推定による結果を統合し，推定した表情とその強度を出力する．

2.3 3 次元モジュール群

・　顔検出・追跡、顔向き推定モジュール

3 次元顔向き推定モジュールは，奥行きセンサから得られる奥行き映像とカメラ映像を用いてユーザの顔領域を検出・追跡し，3 次元実空間で顔向きを推定する．具体的には汎用奥行きセンサの Kinect ライブラリを活用し，顔向き推定に有用な情報を取得する[4].

本モジュールはまずカメラ映像からユーザの顔領域を検出し，隣接フレーム間の対応を考慮しながら顔領域内の 87 の特徴点を検出する．これらの特徴点を奥行き映像に投影し，顔の 3 次元モデルを当てはめながら，最適な特徴点の位置を調整する．本モジュールから 3 次元実空間上の頭の傾き（yaw, pitch, roll）が出力される．

・　表情推定モジュール

表情推定モジュールは，Action Unit 特徴量から TV 視聴者の表情変化を検出する．Action Unit (AU)とは顔の筋肉の動作単位であり，各 AU 値を用いることで，表情変化に伴う顔の動きを客観的に記述できる．汎用奥行きセンサの Kinect ライブラリから 6 種類の AU 値をフレーム毎に取得し，その特徴量を用いて表情表出検出器を学習した．

3. 実験

3.1 実験条件

提案システムの有効性を確認するため，家庭を模した環境で実験を行った．一般的なリビング空間を疑似的に作成し，30 人の一般被験者を招いて TV 視聴状況を計測した．日常生活では TV は家事や仕事をしながら，いわゆる「ながら視聴」されることが多い．そのため，実験中は被験者の行動に一切制限を設けず，PC・携帯操作や部屋の移動を許容し，普段通りの振る舞いを依頼した．図 2 に実験中の被験者の行動例を示す．

一人あたり 2 時間の実験を行い，その中で 15 番組（各 5～15 分）を再生した．歴史，音楽，旅行，教育，バラエティなど多岐にわたるジャンルの映像を再生し，被験者の興味が偏らないよう配慮した．

また被験者を収録した映像を第三者が確認し，1 秒単位で注視状態（見ている/見ていない），および表情表出（表情の有無）をアノテーションした．このアノテーションデータを正解データとし，注視状態検出器および表情表出検出器を作成した．

図 2　TV 視聴実験での被験者の行動例

3.2 注視状態推定精度

まず，注視状態の推定精度を評価した．注視状態判定器は 29 人分のデータを学習，残りの 1 人のデータをテストに使い，いわゆる 30-fold 交差検定で評価した．学習では，第三者が目視アノテーションした注視正解データを基に，RBF カーネルの 2 値 SVM 識別器を作成した．

識別結果を図 3 に示す．常に注視，もしくは非注視と出力する“固定出力”，“ランダム出力”，および“顔検出”よる結果をベースラインとした．アノテーションデータから算出した注視時間率の割合は 47.6%であり、この値が注視固定出力の精度と一致する（非注視固定出力は 52.4%）．顔検出ベースラインでは，一般的な Viola-Jones の手法で顔検出が成功したフレームを注視と判定した[5]．顔の検出は注視と高い相関があると予想されたが，その精度は 54.2%にとどまった．

提案手法の精度は各ベースラインの精度に対して 30%程高い 80.1%であり，本手法の有効性を確認できた．2 次元、3 次元の顔向き推定モジュールが両者の長所を活かしながら頑健に顔追跡している点が，高い注視状態推定精度に貢献した．

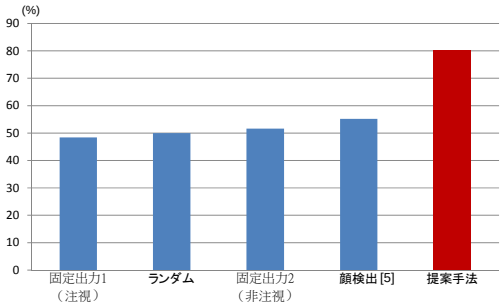

図 3　注視状態推定精度

3.2 表情表出推定精度

続いて表情表出推定精度を検証した．TV 視聴実験データから約 20,000 フレームの表情表出画像を抽出し，同数の無表情画像を加えて検証に用いた．表情表出を判定する 2 値 SVM を被験者ごとに作成し，2-fold 交差検定で評価した．

作成した表情表出検出器で表情の有無を検出したところ，平均適合率は 70.0%であり，表情表出の有無を頑健に判定できることを確認した．

3. まとめ

家庭用映像センサのデータを利用し，TV 視聴者の顔向きと表情表出を推定するシステムを開発した．センサから得られるデータを解析し，顔領域の検出・追跡，顔向き推定，表情表出を行う機能をモジュール化するとともに，モジュールを 2 次元群，3 次元群に分割し，並行して処理を行うことで頑健性を高めた．実生活空間を模した環境での TV 視聴実験を通し，家庭での自由な振る舞いに対しても本システムが高い精度で TV への注視状態と表情表出を検出できることを確認した．注視状態と表情表出結果から視聴者個々の番組評価を推定することが可能であり，本システムがユーザ個人の趣味・嗜好の理解に有効であることを確認した．

謝　辞

本研究の一部は，総務省の委託研究「生活空間における人の注視に着目した映像コンテンツ評価手法に関する研究開発」として実施したものです．

文　献

[1] Microsoft, USA. XBOX Kinect, DOI=http://www.xbox.com/kinect

[2] S. Clippingdale and M. Fujii, “Video face tracking and recognition with skin region extraction and deformable template matching,” International Journal of Multimedia Data Engineering and Management, vol. 3, no.1, pp.36-48, 2012.

[3] 奥田誠，藤井真人，佐藤洋一, “表情強度推定値と Bayes 推定を組み合わせた顔表情認識”, 情報科学技術フォーラム講演論文集（FIT 2013）, no.3, J-028, p.439-440, 2013.

[4] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang, “3D deformable face tracking with a commodity depth camera,” In proc. of the European Conference on Computer Vision (ECCV), pp. 229-249, 2010.

[5] P. Viola and M. Jones: Rapid object detection using a boosted cascade of simple features, In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511–518 (2001)